EVALUATION OF A NEW SPECTRAL PEAK CODING STRATEGY FOR THE NUCLEUS 22 CHANNEL COCHLEAR IMPLANT SYSTEM


ABSTRACT

Sixty-three postlinguistically deaf adults from four English-speaking countries participated in a 17-week field study of performance with a new speech coding strategy, Spectral Peak (SPEAK), and the most widely used strategy, Multipeak (MPEAK), both of which are implemented on wearable speech processors of the Nucleus 22 Channel Cochlear Implant System; MPEAK is a feature-extraction strategy, whereas SPEAK is a filterbank strategy. Subjects' performance was evaluated with an experimental design in which use of each strategy was reversed and replicated (ABAB). Average scores for speech tests presented sound-only at 70 dB SPL were higher with the SPEAK strategy than with the MPEAK strategy. For tests in quiet, mean scores for medial vowels were 74.8 percent versus 70.1 percent; for medial consonants, 68.6 percent versus 56.6 percent; for monosyllabic words, 33.8 percent versus 24.6 percent; and for sentences, 77.5 percent versus 67.4 percent. For tests in noise, mean scores for Four-Choice Spondees at +10 and +5 dB signal-to-noise ratio (S/N) were 88.5 percent versus 73.6 percent and 80.1 percent versus 62.3 percent, respectively; and for sentences at +15 dB, +10, and +5 dB S/N, 66.5 percent versus 43.4 percent, 61.5 percent versus 37.1 percent, and 60.4 percent versus 31.7 percent, respectively. Subjects showed marked improvement in recognition of sentences in noise with the new SPEAK filterbank strategy. These results agree closely with subjects' responses to a questionnaire on which approximately 80 percent reported they heard best with the SPEAK strategy for everyday listening situations.

The Nucleus multielectrode intracochlear implant system was first implanted in 1982; since then it has been implanted in over 8000 recipients. For over a decade, there has been ongoing development of speech coding strategies at the University of Melbourne to improve patients' ability to perceive temporal and spectral cues of conversational speech. Major work has focused on strategies that extract, from incoming sound, features important for speech recognition. With the first speech coding strategy,

*Department of Otolaryngology, Washington University, St. Louis, Missouri
†Cooperative Research Centre, Department of Otolaryngology, University of Melbourne, East Melbourne, Victoria, Australia
‡Cochlear Proprietary Limited, Lane Cove, New South Wales, Australia
§Cochlear Corporation, Englewood, Colorado
*Sunnybrook Health Science Center, University of Toronto, Toronto, Ontario, Canada
**Denver Ear Institute, Denver, Colorado
††University of Sydney Cochlear Implant Clinic, Sydney, New South Wales, Australia
§§St. Paul's Hospital, Vancouver, British Columbia, Canada
###Michigan Ear Institute, Farmington Hills, Michigan
####South of England Cochlear Implant Centre, Institute of Sound and Vibration Research, University of Southampton, Southampton, United Kingdom

Portions of this manuscript were presented at the Annual Meeting of the American Otological Society, Palm Beach, Florida, May 7-8, 1994.

Research was funded by the Cochlear Proprietary Limited, Cochlear Corporation, Cochlear A.G., and the participating implant centers.

Preparation of the manuscript was supported by grant R01·00581 from the National Institute on Deafness and Other Communication Disorders.

Reprint requests: Dr. Margaret W. Skinner, Department of Otolaryngology—Head and Neck Surgery, Washington University School of Medicine, 517 South Euclid Avenue, Campus Box 8115, St. Louis, MO 63110
F0F2, the voice fundamental frequency (F0) as well as the amplitude and frequency of peak sound energy in the region of the second formant (F2) of speech (800–2300 Hz) were extracted. The fundamental frequency determined the stimulation rate, and F2 information determined the electrode to be stimulated and amplitude of stimulation. The F2 frequency range was divided into logarithmically equal passbands that were assigned to active electrodes in an apical-to-basal order. With this assignment of frequency bands, stimulation of the F2 electrode caused pitch percepts according to the tonotopic organization of the cochlea. The second speech coding strategy, F0F1F2, added information about the first formant (F1) of speech (300–1000 Hz), and the bassward stimulation of second formant information (F2) was changed slightly (1000–4000 Hz). With this strategy, F1 frequency bands were divided in equal linear steps and assigned to the apical third of the electrodes, and F2 frequency bands were divided in equal logarithmic steps and assigned to the basal two-thirds of the electrodes. The third speech coding strategy, Multipeak (MPEAK), added information about the amplitude in three high-frequency bands (2000–2800 Hz, 2800–4000 Hz, and 4000–6000 Hz) to the F1 and F2 information. Amplitude information in the high-frequency bands was delivered to three fixed basal electrodes. With the evolution of speech coding strategies, progressively greater numbers of stimulation pulses per analysis cycle (F0F2 = 1, F0F1F2 = 2, MPEAK = 4) have been presented in rapid succession. With each increase in the number of electrodes stimulated, there has been an increase in temporal and spectral information delivered to the auditory system and a concomitant improvement in patients’ speech recognition by sound alone. Although most implant patients use the MPEAK strategy, a few patients use the F0F2 or F0F1F2 strategy. The Mini Speech Processor (MSP), which has been in clinical use since 1989, can be programmed for any of these strategies.

In 1989, a wearable speech processor called the Spectral Maxima Sound Processor (SMSP) was developed at the University of Melbourne, first for research use with an advanced multichannel cochlear implant developed by the University and then modified for research with the Nucleus implant. The speech coding strategy used with the SMSP processor was fundamentally different from that used in the previous strategies, which were based on feature extraction. With this strategy, incoming sound was analyzed using 16 digital bandpass filters, each of which was assigned in tonotopic order to 16 of the user’s implant electrodes, with the lowest frequency band assigned to the most apical electrode used by that subject. For each analysis cycle, the microprocessor scanned the 16 spectral estimates, selected the six channels with largest amplitudes, or “maxima,” and presented these maxima sequentially from largest to smallest on the associated electrodes. If spectral energy was detected in fewer than six channels, fewer electrodes were stimulated. The average length of an analysis cycle was 4 ms.

There were several reasons why the filterbank approach of the Spectral Maxima speech coding strategy provided more information than the MPEAK feature extraction strategy. With the Spectral Maxima strategy, any six of the 16 electrodes were stimulated as the maxima of incoming sound shifted across the frequency range of the processor. In most cases, these maxima represented speech formants, but the Spectral Maxima strategy provided better spectral detail, owing to stimulation of adjacent channels rather than estimation of just one “peak” for each formant. In addition, there was rapid sequential stimulation of six electrodes with the Spectral Maxima strategy instead of four electrodes with the MPEAK strategy for each analysis cycle.

Performance of patients using the Spectral Maxima strategy in research at the University of Melbourne was encouraging. In the first study, the four subjects performed significantly better on sound-only speech tests when using this strategy than when using the MPEAK strategy. Average improvement was 15 percent for closed-set vowels, 15.5 percent for closed-set consonants, 17.5 percent for open-set monosyllabic words, and 28.7 percent for sentences in noise. Further analysis of subjects’ responses on the vowel, consonant, and word tests using the two speech coding strategies revealed what aspects of speech feature perception contributed to this improvement. For the monosyllabic words, there was significant improvement on all features for vowels (duration, F1, F2, F1-transition, F2-transition) and all features for consonants (voicing, place, and manner of articulation) with the Spectral Maxima strategy. For closed-set vowel and consonant tests, there was significant improvement for vowel F1 and F2 features, and consonant manner and place of articulation with the Spectral Maxima strategy.

These results prompted development and manufacture of the Spectra 22 processor and Spectral Peak (SPEAK) speech coding strategy from the SMSP research device. The major changes incorporated in the Spectra 22 processor were as follows. A new custom monolithic integrated chip was designed and manufactured. Incoming sound was analyzed with a maximum of 20 digitally programmable analog filters, from which the amplitude in each filter was extracted. The filter outputs were scanned to select an average of six maxima. The actual number of maxima, that ranged between one and ten, was determined by the signal level, spectral composition of incoming sound, and the individual subject’s speech processor program. This scanning occurred continuously, and electrodes associated with the maxima filters were stimulated at the actual scanning rate from a basal-to-apical position in the cochlea. The average rate was 250 Hz, but it could vary by approximately plus or minus 100 Hz, depending on the subject’s speech processor program as well as the incoming sound intensity and spectral composition.
Within an individual subject's program, rate variation was smaller.

The SPEAK speech coding strategy required more power than the MPEAK speech coding strategy because an average of six electrodes were stimulated instead of four. For this reason, a rechargeable battery gave less than the 16 hours use typically obtained with the MPEAK strategy.

The gain at the output of each bandpass filter could be set between values of 1 and 15 in 1-dB increments. The software automatically selected a value of 8 for all filters; however, it could be manually changed to another value for each filter.

The upper and lower frequency boundaries of the bandpass filters were preselected by the software. For a given number of active electrodes, the clinician could choose one from among several sets of bandpass filters. For the maximum number of 20 active electrodes, the software automatically selected 20 bandpass filters for the frequency range from 150 to 10,823 Hz. For 16 electrodes, this selection included the frequency range from 150 to 5744 Hz. Alternative sets of bandpass filters were available that started as low as 75 Hz but could go no higher than 10,823 Hz. The filter bandwidths were linearly distributed from 250 to approximately 1850 Hz and were then logarithmically distributed to the maximum frequency. Scaling of these bandpass filter frequencies made it possible for patients with a range of available electrodes to be programmed.

The noise suppression circuit of the Spectra 22 processor was not designed to subtract the noise floor from the incoming signal, a feature which was included in the MSP noise suppression circuit. Consequently, a base-level increase and enhanced autosensitivity function were incorporated into the Spectra 22 noise suppression circuit. For example, when the processor was set on S, this circuit caused the default base level of 4 to be increased to 10. For this base-level increase, the upper 22.7 dB of incoming sound was processed instead of the usual range of 29.5 dB. In addition, the autosensitivity function of the Spectra 22 was more active than in the MSP. For the MSP set to S, the autosensitivity function reduced the sensitivity level when the noise floor reached 10 dB below the peak level. For the Spectra 22 set to S, this happened when the noise floor reached 12.4 dB below the peak level.

This report describes the results of a 17-week field study comparing performance of 63 postlinguistically deafened adults using the SPEAK strategy implemented on the Spectra 22 processor with performance using the MPEAK strategy implemented on the MSP.

MATERIAL AND METHODS

Field Study Sites and Subjects

Eight cochlear implant centers in four English-speaking countries participated in the field study. Twenty-four adults participated in Australia: 12 at the University of Melbourne Cochlear Implant Clinic at the Royal Victorian Eye & Ear Hospital and 12 at the University of Sydney Cochlear Implant Clinic at the Royal Prince Alfred Hospital. Twenty-three adults participated in the United States: 11 at the Denver Ear Institute, nine at the Department of Otolaryngology of Washington University School of Medicine, and three at the Michigan Ear Institute. Twelve adults participated in Canada: eight at the Sunnybrook Health Science Center of the University of Toronto and four at St. Paul's Hospital in Vancouver. Four adults participated in the United Kingdom at the South of England Cochlear Implant Centre of the Institute of Sound and Vibration at the University of Southampton.

Subjects were 63 postlinguistically deafened adults who were implanted between 20 and 78 years of age (median: 44 yr) who spoke English as their primary language and had used the MSP programmed with the MPEAK speech coding strategy continuously for at least 8 months prior to the study. For 52 subjects, median duration of deafness prior to implantation was 5 years (mode: 1 yr; range: 4 mo–60 yr). Subjects used 16 or more active intracochlear electrodes in their speech processor programs except for six subjects in Melbourne who used between 12 and 15 electrodes. Maximum acceptable loudness levels (MALs) on electrodes in the subjects’ programs averaged 185 or less on the Stimulus Level scale (for a description, see Skinner et al). Subjects who participated represented a wide range of open-set speech recognition on auditory-only sentences in quiet with minimum performance of 5 percent correct on recorded CID Everyday Sentences or Bamford-Kowal-Bench Sentences. This group of subjects met selection criteria that represented the majority (75%) of clinical patients using the Nucleus 22 Channel Cochlear Implant system.

Equipment and Test Environment

An excellent description of the Nucleus multichannel cochlear implant, the Mini 22, its implantation and function is given by Clark et al. The Mini Speech Processor its function and fitting parameters, and the MPEAK speech coding strategy are described by Skinner et al. The SPEAK and MSP processors are the same in the following ways: size and weight; powered with one 1.5 V rechargeable or disposable battery; sensitivity control function; conversion of incoming amplitude to a digital code; range of base levels; range of Q values; availability of Current Level and Stimulus Level amplitude coding strategies; and storage of patient speech processor programs with Random Access Memory.

Each center used two cassette tape recorders connected to an audiometer or laboratory equipment to present recorded speech tests and mix eight-talker babble with the speech at the appropriate
levels and signal-to-noise ratios (S/N). The output of this equipment was delivered to a power amplifier that drove a loudspeaker. A sound-level meter was used to calibrate the stimuli at the center of where the subject's head would be for testing.

Each subject's speech processor was programmed using an IBM-compatible computer, Dual Processor Interface, Version 6.60 of the Diagnostic Programming Software for the MSP and MPEAK strategy, and experimental software (Version 6.90) for the Spectra 22 and SPEAK speech coding strategy.

Subjects were seated between 1 and 2 meters at 0 degrees azimuth to the loudspeaker in a sound-attenuating room or booth that had relatively little reverberation.

Test Materials

Test materials included speech tests and a questionnaire. Speech tests included sound-only recordings of nonsense syllables, monosyllables, and sentences presented in quiet. In addition, sentences and spondees were mixed with eight-talker babble and presented from the same loudspeaker.

The North American vowel set included 14 medial vowels in an /hVd/ context (/i, I, e, æ, a, A, U, u, a, r, ə, ɔ, or, ər, ɜr/); there were four tokens of each vowel in each of nine randomizations. The Australian vowel set included 11 medial vowels in the same context (/i, u, ə, a, ɔ, I, ɔ, æ, æ, a, U, U/); there were four tokens of each vowel in each of eight randomizations. In Southampton, no vowel test was used. The North American consonant set included 14 medial consonants in an /aCa/ context (/b, d, g, p, t, k, f, v, s, z, j, ʒ, m, n, n/); there were five tokens of each consonant in each of the nine randomizations. The Australian and United Kingdom consonant sets included 12 medial consonants (/b, p, m, v, f, d, t, n, z, s, g, k/) in the same context; there were four tokens of each consonant in each randomization. The vowel and consonant tests were presented in a closed-set format in which any of the other phonemes in the test were possible response choices.

The Consonant-Vowel Nucleus-Consonant (CNC) monosyllabic word test consisted of ten lists each with 50 words. An equal proportion of phonemes in spoken American English are included in each list. The CUNY sentence test was developed at City University of New York for audiosvisual research with sensory aids. 132 lists, each containing 12 sentences, were created. Sentence length varies from three to 14 words, with 102 words per list. Lists are scored for total number of correctly recognized words. In the original version, the subject knew the topic of each of the 12 sentences in a list. For the present study, topic words were omitted, and new recordings were made with each list's topic order randomized. Since subjects in Melbourne had heard the CUNY sentences before, a sentence test based on the Speech Intelligibility Test for Deaf Children (SIT) was used.19 This test consists of 40 lists, each containing 15 sen-


tences scored on the basis of 80 key words per list. A comparison of relative difficulty between the CUNY and SIT sentence test recordings for the same subjects suggested that the SIT test was more difficult than the CUNY test.

From the original version of the Four-Choice Spondee test,20 for which there was one list of 20 target words, twelve 20-word lists were created from a larger corpus of words. In this closed-set test, the test word was one of four from which the subject would choose.

Several sets of recordings were made: a North American male talker recorded all speech tests used in the United States and Canada; an Australian male talker recorded the vowel, consonant, and CNC word tests used in Melbourne and Sydney as well as the Four-Choice Spondee test used in Sydney; an Australian female talker recorded the CUNY sentence test used in Sydney and Southampton as well as the CNC word test used in Southampton; another Australian male talker recorded the SIT sentence test used in Melbourne; and a British female talker recorded the consonant test used in Southampton.

The eight-talker babble recording was created by overlaying two identical recordings of four-talker babble (originally recorded by Auditec of St. Louis; rerecorded at the University of Melbourne) with a slight offset in time to create a dense, relatively constant, competing message.

A questionnaire was devised on which subjects were asked to choose the speech processor that provided the best hearing and understanding of speech and other sounds in a number of common listening situations in everyday life.

Procedures

Speech Processor Settings

Prior to the study, each subject's MSP program was fine-tuned to set threshold and MAL for each active electrode so that (1) speech and environmental sounds in everyday life were comfortable and relatively natural sounding when the sensitivity control was set at its optimum, midrange setting, (2) loud speech was not too loud, and (3) soft speech was audible. Sixty-one subjects used the Stimulus Level and two subjects used Current Level for amplitude coding. Choice of Q value, base level, and MPEAK band electrodes was the same as each subject had been using prior to the study. For the SPEAK program, thresholds and MALs from the MSP program were used; occasionally, a small decrease in the thresholds and/or MALs was made to prevent sound being too loud. For subjects who used 21 or 22 electrodes in their MSP programs, one or two of the most basal electrodes were removed from their SPEAK program. Many subjects used a gain of 8 on all bandpass filters. For those who found that the SPEAK strategy provided too much low-frequency sound or that conversational speech sounded fuzzy or mumbly
compared to the MPEAK strategy, the gain was re­duced to 6 on two to four apical electrodes. For those who found the SPEAK strategy provided high-fre­quency sound that was irritating, either the gain was reduced to 6 on two to four basal electrodes or one or more basal electrodes were eliminated from the program.

For 57 subjects, frequency boundaries assigned to electrodes for the SPEAK strategy were the soft­ware-selected set that depended on the number of active electrodes (between 16 and 20). Six subjects used other criteria for assignment of frequency boundaries to electrodes. For subjects who used the software-selected set for the SPEAK strategy, assignment of frequency boundaries to electrodes for the MPEAK and SPEAK strategies was different. The largest difference was for subjects with 22 active elec­trodes in MPEAK and 20 active electrodes in SPEAK. For MPEAK, 1000 Hz was the upper frequency bound on electrode 15, whereas for SPEAK, 1000 Hz was near the lower boundary of frequencies assigned to electrode 18. The smallest difference was for subjects with 16 active electrodes for both strategies. For MPEAK with these subjects, 1000 Hz was the upper frequency bound on electrode 15, whereas for SPEAK, 1000 Hz was near the lower bound of frequencies assigned to electrode 16. If sound quality was unacceptable with this software-selected set, one or two of the most basal electrodes were removed in the SPEAK program. Reduction in number of elec­trodes caused a decrease in the highest frequency analyzed by the filters.

Once MPEAK and SPEAK programs were cre­ated that each subject found acceptable in everyday life, they were not changed for the duration of the study. The optimum sensitivity control setting ob­tained with the Diagnostic Programming Software (Cochlear Pty Ltd., Lane Cove, New South Wales, Australia) and the normal function setting were used for evaluating performance on speech tests.

**Experimental Design**

Speech tests were chosen to evaluate speech rec­ognition at the phoneme, word, and sentence level of linguistic complexity. As had been shown in previous research,1,2 use of sentences in noise was particularly important for evaluating performance differences with the two speech coding strategies. Sentence recordings were chosen so that most sub­jects could recognize at least a few words at +10 dB S/N with the MPEAK strategy. The second S/N ratio was chosen so that meaningful data could be ob­tained for each subject; that is, the scores would be above the chance level and below approximately 90 percent correct. A +5 dB S/N was chosen for subjects who had 50 percent or better scores at +10 dB S/N during a practice session prior to the first evaluation with MPEAK, a +15 dB S/N was chosen for those with scores of less than 50 percent at +10 dB S/N. For subjects whose scores on the CUNY sentence test in quiet were 35 percent or less, the easier, Four-Choice Spondee, test was presented. The CUNY or SIT sentence test was presented in quiet as a baseline for performance in noise.

The number of weeks subjects used the MPEAK and SPEAK speech coding strategies was chosen as a compromise between sufficient time to listen with each strategy and a reasonable length of time for the study. Since subjects had used the MPEAK strategy for at least 8 months, they used this strategy for 3 weeks during the first time period (A1). Because the SPEAK strategy had never been used before, subjects used this strategy for 6 weeks during the first time period (B1). When they returned to the MPEAK and SPEAK strategies for the second time period (A2, B2), subjects used each strategy for 3 weeks. The use of each strategy at two time periods made it possible to determine whether learning affected performance.

It was necessary and appropriate to use a single-subject design because of the heterogeneous nature of profoundly deaf subjects and expected variability in outcomes. Because subjects had a wide range of auditory capabilities and were accustomed to different English dialects, it was important to select speech materials and signal-to-noise ratios that would yield meaningful results for each subject. Although subjects took the same types of tests, differences in test materials were as follows: number of vowels and con­sonants; sentence materials (CUNY and SIT tests); Four-Choice Spondee test in noise for five subjects with limited open-set speech recognition; use of +15 or +5 dB S/N for the second condition of sentences in noise; speech test recordings of talkers with different dialects and gender. For several tests, the number of subjects in the database was less than 63. The following results were not included in the database: (1) vowel and consonant tests for four subjects in the United Kingdom, (2) CNC phoneme scores for Aus­tralian subjects, and (3) CUNY sentences in quiet for two Sydney subjects who listened to the Four-Choice Spondee test in noise. With the single-subject design, these differences in the test protocol did not adversely affect interpretation of the results.

The test protocol was designed to evaluate each subject’s test-retest variability on each speech test for each speech coding strategy and time period. For all except the sentence tests, one list was presented at each of two evaluation sessions for each strategy and time period. Since there is believed to be more list-to-list variability in test difficulty for the sentence tests, two lists per condition were presented at each session. This evaluation of test-retest variability provided data necessary for evaluation of the difference in performance associated with the two speech coding strategies and possible learning during the study.

**Evaluation Schedule**

Two-hour sessions on weeks 1 and 2 were devoted to fine-tuning the MPEAK program, practice taking the same tests but with different lists than were used...
later for the evaluation, and deciding whether to present CUNY sentences at +15 or +5 dB S/N or to use the Four-Choice Spondees in noise. Sessions on weeks 3 and 4 were devoted to a test-retest evaluation with the MPEAK strategy using one list each of the vowels, consonants, and CNC words, two lists of sentences in quiet, and either two lists of sentences (58 subjects) or Four-Choice Spondees (5 subjects) at each of the two signal-to-noise ratios (+10 and +5 or +15 dB) at each session. There were three exceptions to this protocol: two Sydney patients who listened to Four-Choice Spondees in noise did not listen to sentences in quiet; Southampton subjects did not listen to a vowel test; and Southampton subjects listened to the consonant test only once instead of twice.

At the end of session 4, subjects started to use the Spectra 22 processor and speech coding strategy, and during sessions in weeks 5, 8, and 9, the SPEAK program was optimized. Further optimization of the SPEAK program was done during sessions in weeks 6 and 7 for some subjects who had difficulty adjusting to the SPEAK strategy. At sessions in weeks 5, 8, and 9, subjects participated in live-voice speech tracking training as well as responded to the practice lists for each speech test. Sessions in weeks 10 and 11 were devoted to evaluation with the SPEAK strategy, using the same protocol and speech tests as with the MPEAK strategy but different randomizations or lists.

Subjects started to use the MPEAK strategy again at the end of the evaluation in week 11. They came in again for evaluation with the MPEAK strategy in weeks 13 and 14. They started to use the SPEAK strategy at the end of the evaluation in week 14. The final evaluation with the SPEAK strategy was in weeks 16 and 17, and they responded to the questionnaire in week 17.

Tests were always given in the same order: vowels, consonants, words, sentences in quiet, sentences at +10 dB S/N, and sentences at +15 or +5 dB S/N (or spondees at +10 and then +5 dB S/N). No list or randomization was given twice during the study (except for the SIT sentences), and the order of lists/randomizations was pseudorandomized within and across subjects. For the Melbourne subjects, the eight SIT sentence lists that were repeated were balanced across subjects and the two strategies during weeks 13, 14, 16, and 17.

In summary, the experiment was a replicated single-subject reversal design (ABAB). Test-retest data for each test was obtained from each subject at each evaluation period, except for the consonant test presented to the Southampton subjects.

### Analysis of Data

Consistent with the experimental design, speech perception data were analyzed for each of the field trial participants. For each subject, separate two-by-two Analyses of Variance (ANOVA) were conducted for each of the six primary dependent measures: medial vowels, medial consonants, monosyllabic words (scored both for words and phonemes), sentences in quiet and sentences at each of two signal-to-noise ratios (or Four-Choice Spondees at each of 2 signal-to-noise ratios). The factors analyzed were the two speech coding strategies (MPEAK and SPEAK) and the two time periods [initial exposure (A1B1) versus replication (A2B2)]. A full factorial model was used to determine main effects and interactions. The decision to evaluate both speech coding strategy and learning factors with ANOVA is a more conservative approach than analyzing only the speech coding strategy factor with the ANOVA or binomial model.

Subjects’ mean scores for each speech coding strategy (Results section) are an average across all lists presented during the four evaluation sessions, two for each time period (i.e., A1 plus A2, and B1 plus B2).

### RESULTS

#### Effects of Speech Coding Strategy on Speech Test Performance

**Medial Vowel Test**

The mean score on the medial vowel test using the MPEAK and SPEAK speech coding strategies is shown for each of the 59 subjects in Figure 1. Mean scores ranged from 29 to 98 percent with the MPEAK strategy and from 36 to 100 percent with the SPEAK strategy; scores exceeded 90 percent for five subjects with MPEAK and 12 subjects with SPEAK. A greater proportion of Australian subjects (9/24) than North American subjects (3/35) scored above 90 percent with the SPEAK strategy. The North American test might have been more difficult than the Australian test because of differences in the number of vowels and/or intelligibility of the talkers. This potential ceiling effect made it difficult to determine whether there was a significant difference in performance between the two strategies for a number of subjects. Fifteen subjects (25.4%) had significantly higher mean scores using the SPEAK strategy, and one subject (1.7%) had a significantly higher mean score using the MPEAK strategy. Across subjects, the mean score using the SPEAK strategy was 5 percent higher than that using the MPEAK strategy.

**Medial Consonant Test**

The mean score on the medial consonant test using the MPEAK and SPEAK strategies is shown for each of the 59 subjects in Figure 2. Mean scores ranged from 19 to 86 percent with the MPEAK strategy and from 28 to 97 percent with the SPEAK strategy. Thirty-three subjects (55.9%) had significantly higher mean scores using the SPEAK strategy than the MPEAK strategy, and none had a significantly higher score with the MPEAK strategy. Across sub-
jects, the mean score using the SPEAK strategy was 12 percent higher than that using the MPEAK strategy. This improvement with the SPEAK strategy is approximately double that found with the vowel test.

**CNC Word Test**

The mean score on the CNC monosyllabic word test (scored according to words) using the two speech coding strategies is shown for each of the 63 subjects in Figure 3. Mean scores ranged from 0.5 to 72 percent with the MPEAK strategy and from 0.5 to 76 percent with the SPEAK strategy. Twenty subjects (31.8%) had significantly higher mean scores using the SPEAK strategy, and one subject (1.6%) had a significantly higher mean score using the MPEAK strategy. Across subjects, the mean score using the SPEAK strategy was 9 percent higher than that using the MPEAK strategy.

The CNC words were scored according to phonemes correct for all except the Australian subjects. The mean score using the two speech coding strategies is shown for each of 39 subjects in Figure 4. Mean scores ranged from 23 to 80 percent with the MPEAK strategy and from 27 to 83 percent with the SPEAK strategy. Sixteen subjects (41.0%) had significantly higher mean scores using the SPEAK strategy, and none had significantly higher scores with the MPEAK strategy. Across subjects, the mean score using the SPEAK strategy was 10 percent higher than that using the MPEAK strategy.
**CUNY and SIT Sentence Tests**

The mean score for sentences in quiet using each of the two strategies is shown for each of the 61 subjects in Figure 5. Although there were a few subjects with scores under 20 percent correct, 13 subjects in MPEAK and 28 subjects in SPEAK had scores that were over 90 percent correct. This probable ceiling effect made it difficult to determine whether there was a significant difference in performance between the two strategies for these subjects. The mean improvement in score of 10 percent across subjects with the SPEAK strategy was limited by this probable ceiling effect. Twenty-four subjects (39.3%) had significantly higher scores with the SPEAK strategy, and two subjects (3.3%) had significantly higher scores with the MPEAK strategy.

Fifty-eight of the 63 subjects responded to sentence tests at +10 dB S/N; their mean scores with the two strategies are shown in Figure 6. The marked improvement in performance with the SPEAK strategy compared with the MPEAK strategy is readily apparent. Forty-one subjects (70.7%) performed significantly better with the SPEAK strategy, and none performed significantly better with the MPEAK strategy. The mean improvement in score across subjects with the SPEAK strategy was 24 percent. This improvement is double that for the consonant test and more than double that for the vowel, CNC word, and sentences in quiet tests.
Forty of the 58 subjects took sentence tests at +15 dB S/N; their mean scores with the two strategies are shown in Figure 7. Thirty-two (80.0%) had significantly higher scores with the SPEAK strategy, and none had significantly higher scores with the MPEAK strategy. Mean improvement in score across subjects with the SPEAK strategy was 23 percent. This improvement is similar to that for sentences at the +10 dB S/N.

Eighteen of the 58 subjects took sentence tests at +5 dB S/N; their mean scores with the two strategies are shown in Figure 8. Fourteen (77.8%) had significantly higher scores with the SPEAK strategy, and none had significantly higher scores with the MPEAK strategy. Mean improvement in score across subjects with the SPEAK strategy was 29 percent, which was slightly higher than for sentences at +10 and +15 dB S/N.

Four-Choice Spondees in Noise

Five of the 63 subjects responded to the Four-Choice Spondee test in noise. The mean score for the two strategies at +10 and +5 dB S/N is shown for each subject in Figure 9. At +10 dB S/N, four subjects (80%) performed significantly better with the SPEAK strategy than with the MPEAK strategy; a ceiling effect may have limited the fifth subject's performance with the SPEAK strategy. Across subjects, the mean score with the SPEAK strategy was 15 percent higher than with the MPEAK strategy.

Figure 5. Mean score (percent correct) on CUNY or SIT sentences in quiet using the MPEAK and SPEAK speech coding strategies for each of 61 subjects. *Significantly different scores.

Figure 6. Mean score (percent correct) on CUNY or SIT sentences presented at +10 dB S/N in eight-talker babble using the MPEAK and SPEAK speech coding strategies for each of 58 subjects. *Significantly different scores.
At the +5 dB S/N, two subjects (40%) performed significantly better with the SPEAK strategy. Since the mean improvement in score (18%) with the SPEAK strategy for +5 dB S/N was slightly greater than that (15%) for +10 dB S/N, it appears that increased variability in scores among lists for the +5 dB S/N condition made it impossible to show a significant improvement for the other three subjects at the +5 dB S/N.

Effects of Time Period and Its Interaction with Speech Coding Strategy on Speech Test Performance

Each subject's scores across both speech coding strategies were compared for the first and second time periods with ANOVA to determine whether there was a significant ($p < .05$) increase or decrease in score as a function of time. The percentage of subjects who showed a significant change in score is shown for each test in Table 1. All subjects obtained significantly higher mean scores during the second half of the study; that is, learning occurred. The subjects' patterns of change in score as a function of time period and speech coding strategy have been categorized and shown in Table 1. In the first pattern (1), use of the SPEAK strategy enabled subjects to obtain a higher mean score with the MPEAK strategy during the second time period ($A_2$) than during the first time period ($A_1$). That is, these subjects understood speech better with the MPEAK strategy after they had listened for 6 weeks to the increased information the SPEAK strategy provided. As shown in...
Table 1, almost all subjects' scores followed this pattern of learning. In the second pattern (2), subjects' scores with the MPEAK strategy during the second time period (A2) decreased after using the SPEAK strategy. For these few subjects, it appeared that they needed longer than 2 or 3 weeks to adjust to the MPEAK strategy. In the third pattern (3), subjects' scores with the MPEAK strategy during the second time period stayed the same as they had been during the first time period, but their scores with the SPEAK strategy increased. In the fourth pattern (4), one subject's mean scores for the vowel test were higher with the MPEAK strategy across both time periods than with the SPEAK strategy although use of the

SPEAK strategy seemed to enhance her scores with the MPEAK strategy during the second half of the study.

For a small percentage of the subjects, there was a significant interaction (p < .05) between the time period and speech coding strategy as shown for each test in Table 2. The same pattern of scores as described for Table 1 was used for categorizing the interaction results. For most subjects, use of the SPEAK strategy resulted in a higher score with the MPEAK strategy during the second time period (pattern 1). Since only a small percentage of subjects showed a significant interaction between the two main effects, this result supports the robustness of the significant improvement provided by the SPEAK strategy compared with the MPEAK strategy.

Table 1. Effects of Time Period on Speech Test Performance

<table>
<thead>
<tr>
<th>Table 1. Effects of Time Period on Speech Test Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Percentage of subjects who had a significantly higher (p &lt; .05) mean score for each speech test across the MPEAK and SPEAK speech processing strategies during the second half of the study than the first half of the study, and whose mean scores for each time period (A1, B1, A2, and B2) followed one of four patterns described in the text.</td>
</tr>
</tbody>
</table>

Table 2. Effects of Interaction of Speech Coding Strategies between Time Periods

<table>
<thead>
<tr>
<th>Table 2. Effects of Interaction of Speech Coding Strategies between Time Periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Percentage of subjects who had significant (p &lt; .05) interaction between the two main effects, time period and speech coding strategy, and whose mean scores for each time period (A1, B1, A2, and B2) followed one of four patterns described in the text.</td>
</tr>
</tbody>
</table>
Table 3. Responses to Questionnaire

<table>
<thead>
<tr>
<th>Listening Condition</th>
<th>Better with</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPEAK</td>
<td>MPEAK</td>
<td>Same</td>
<td>Don't Know</td>
</tr>
<tr>
<td>Understanding speech in quiet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two friends talking nearby</td>
<td>81.4</td>
<td>0.0</td>
<td>16.9</td>
<td>1.7</td>
</tr>
<tr>
<td>Child talking (6-10 yr)</td>
<td>66.1</td>
<td>3.6</td>
<td>14.3</td>
<td>16.1</td>
</tr>
<tr>
<td>Female voices</td>
<td>82.8</td>
<td>3.4</td>
<td>13.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Male voices</td>
<td>75.9</td>
<td>5.2</td>
<td>19.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Friend talking outdoors</td>
<td>82.5</td>
<td>0.0</td>
<td>14.0</td>
<td>3.5</td>
</tr>
<tr>
<td>Relative talking at normal level</td>
<td>81.0</td>
<td>0.0</td>
<td>13.8</td>
<td>5.2</td>
</tr>
<tr>
<td>Telephone</td>
<td>75.0</td>
<td>3.6</td>
<td>12.5</td>
<td>8.9</td>
</tr>
<tr>
<td>TV News</td>
<td>87.7</td>
<td>1.8</td>
<td>8.8</td>
<td>1.8</td>
</tr>
<tr>
<td>TV Movies</td>
<td>82.5</td>
<td>0.0</td>
<td>14.0</td>
<td>3.5</td>
</tr>
<tr>
<td>Radio</td>
<td>66.7</td>
<td>0.0</td>
<td>22.2</td>
<td>11.1</td>
</tr>
<tr>
<td>Understanding speech in noise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech</td>
<td>82.8</td>
<td>1.7</td>
<td>8.6</td>
<td>6.9</td>
</tr>
<tr>
<td>Friends at social gathering</td>
<td>84.5</td>
<td>0.0</td>
<td>10.3</td>
<td>5.2</td>
</tr>
<tr>
<td>Conversation in the car</td>
<td>82.8</td>
<td>1.7</td>
<td>12.1</td>
<td>3.4</td>
</tr>
<tr>
<td>Conversation at dinner table</td>
<td>79.7</td>
<td>1.7</td>
<td>13.6</td>
<td>5.1</td>
</tr>
<tr>
<td>Listening at home</td>
<td>87.7</td>
<td>1.8</td>
<td>8.8</td>
<td>1.8</td>
</tr>
<tr>
<td>Listening at work</td>
<td>87.2</td>
<td>2.1</td>
<td>8.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Environmental sounds</td>
<td>81.4</td>
<td>0.0</td>
<td>15.3</td>
<td>3.4</td>
</tr>
<tr>
<td>Music</td>
<td>83.9</td>
<td>0.0</td>
<td>10.7</td>
<td>5.4</td>
</tr>
</tbody>
</table>

*Percentage of subjects responding that they heard better with the MPEAK or SPEAK speech coding strategy, heard the same with both strategies, or didn’t know with which they heard best for each listening condition.

Responses to Questionnaire

A summary of subjects' responses to the questionnaire concerning listening situations in everyday life is shown in Table 3. These responses reflect the dramatic improvement that use of the Spectra 22 processor and SPEAK speech coding strategy made in all situations. Over 80 percent of the subjects said that the SPEAK strategy enabled them to hear and understand better than with the MPEAK strategy in 13 of the 18 listening situations, whereas few (0.0-5.2% of the subjects; mode: 1.7%) said that the MPEAK strategy was better. Almost all subjects reported that they heard as well or better with the SPEAK strategy in everyday life than with the MPEAK strategy.

DISCUSSION

There was a significant improvement in speech recognition for 60 of the 63 subjects on at least one test when using the SPEAK speech coding strategy. More important, there was marked improvement in subjects' scores for sentences in noise, with some subjects obtaining as much as a 200-300 percent increase in score with the SPEAK strategy compared with the MPEAK strategy. In addition, no subject obtained a significantly lower score with the SPEAK strategy for the sentence and Four-Choice Spondee tests presented in noise. Although the mean improvement across subjects on medial vowels, medial consonants, CNC monosyllabic words, and sentences in quiet was more modest, some subjects showed marked improvement in performance and only one or two subjects had a significant decrement in performance with the SPEAK strategy on isolated tests. It is possible that their performance with the SPEAK strategy would have been better if they had a longer time to use the SPEAK strategy.

Subjects' significant improvement in performance with the SPEAK strategy on speech tests presented at a slightly raised vocal effort (70 dB SPL) in the laboratory is strongly supported by their reported benefit with the SPEAK strategy in many common listening situations in everyday life (see Table 1). The generality of these findings is enhanced by the inclusion of a large number of postlinguistically deafened subjects with a wide range of open-set speech recognition abilities from eight centers in four English-speaking countries in a study with a single research design. This research design included measures of test-retest variability and replication of evaluation of the MPEAK and SPEAK speech coding strategies that made it possible to determine the effect of learning and its interaction with speech coding strategy on the performance of each subject. As shown in Table 2, between 6.8 and 40 percent of the subjects who responded to the ten speech tests had significant learning effects. However, the performance of only a small percentage of subjects showed a significant interaction between this learning effect and the speech coding strategy effect. Use of Analysis of Variance with a full factorial model (2 strategies and 2 time periods) to evaluate the results
for each subject for each speech test made it possible to provide a conservative estimate of the significant improvement subjects obtained with the SPEAK strategy and to demonstrate the robustness of this improvement, even though there was a learning effect during the study for some subjects.

Although the average improvement on vowels, consonants, and CNC words with the SPEAK compared to the MPEAK strategy was less in this study than in a study with four subjects using the SMSP (instead of SPEAK), the average performance for sentences in noise for the two studies is comparable. In the present study, 65 instead of four subjects participated, and they had a much wider range of auditory capabilities than the four subjects in the SMSP study. The generality of the present study’s results lends strong support for the initial findings with the SMSP.

In summary, the performance and benefit reported by the 63 postlingually deafened subjects who participated in this study indicate that the Spectra 22 processor and SPEAK speech coding strategy enable almost all subjects to recognize speech as well or significantly better than with the MPEAK strategy that they had used for at least 8 months prior to this study.

REFERENCES


We are grateful to the 63 subjects who have contributed so much to this research. We thank the following colleagues who contributed to the manufacture of the SPEAK processor, development of the experimental software, trouble-shooting of the equipment, analysis of data, and data collection, and offered critical comments on an earlier draft of this paper: Jim Patrick, Andrew Mortlock, Kevin Walls, Marilyn Demorest, Timothy Holden, Laura Holden, Theresa Brunelli, Kristine Ash, Rachel Smith, Marissa Skok, Kerrie Plant, Elvira Gerin, and Rodney Hollow.
Author/s:
Skinner, Margaret W.; Clark, Graeme M.; Whitford, Lesley A.; Seligman, Peter M.; Staller, Steven J.; Shipp, David B.; Shallup, Jon K.; Everingham, Colleen; Menapace, Christine M.; Arndt, Patti L.; Antogenelli, Trisha; Brimacombe, Judith A.; Pijl, Sipke; Daniels, Paulette; George, Catherine R.; McDermott, Hugh J.; Beiter, Anne L.

Title:
Evaluation of a new Spectral Peak coding strategy for the Nucleus 22 channel cochlear implant system

Date:
1994

Citation:

Persistent Link:
http://hdl.handle.net/11343/27401